

**Neighborhood Effects and Trial on the Internet:
Evidence from Online Grocery Retailing**

David R. Bell and Sangyoung Song

Grocery Store Antitrust:
Historical Retrospective and Current Developments

May 24, 2007

Outline

- Background
 - Motivation and definitions
- Research Questions and Approach
- Model
 - Elements and assumptions
 - Neighborhood effects
- Empirical Analysis
 - Data
 - Preliminaries
 - Results
- Conclusion

Background

“... choice of a store location has a profound effect ... A bad choice may all but guarantee failure, a good choice, success.”

“Store Location: Little Things Mean A Lot” *CBSC*.

For the Internet retailer, however ...

- Geographical boundary of customer base constrained only by availability of shipping infrastructure
- Multiplicity of competitors and customer options

⇒ Relevance of interaction?

⇒ Implications for evolution?

Background

Empirical distribution of (a) revenue and (b) average value reveals

- California, Texas, Florida, New York generate most revenue
- Interior western states have larger orders
- *Observational units*: Individuals denoted by i reside in regions denoted by z , may experience an event at t
- *Social contagion/neighborhood effect*: Local spillovers resulting from (observational) learning and/or direct communication
- *Risk set*: Group of regions that have yet to experience the event (regions); sum of all individuals over all regions (individuals)

Background

- Agents' decisions influenced by decisions of others
 - BANERJEE (1992)
 - BIKHCHANDANI, HIRSHLEIFER & WELCH (1992)
- Economic processes may generate spatial patterns
 - CASE, HINES & ROSEN (1993)
 - GOOLSBEE & KLENOW (2002)
- Social networks disseminate information
 - OYEN & DE FLEUR (1953)
 - TOLENAY, DEANE & BECK (1996)
- External information more relevant for “trial”
 - HOWARD & SHETH (1969)
 - URBAN (1975)

Research Questions and Approach

1. Are neighborhood effects present in trial of Internet service?
 - (a) Is the effect identifiable and consistent with rational behavior?
 - (b) If present, what is the economic impact on space-time diffusion?
 - (c) (Are neighborhood effects absent for repeat?)

⇒ Approach

- Link statistical theory of hazard to random utility
- Estimate effect with appropriate controls

Model

Elements and assumptions

- Instantaneous probability that event occurs for individual i at time t

$$\lambda(t) = \lim_{\Delta \rightarrow 0} P(t \leq T_{iz} \leq t + \Delta | T_{iz} \geq t) / \Delta$$

- The discrete time analog is

$$P_{iz}(t) = P(T_{iz} = t | T_{iz} \geq t, X_{iz}(t))$$

- T_{iz} is a discrete random variable denoting uncensored time of trial
 - Expression is also a conditional probability
- Individual i at location z has an unobserved utility value for trial at t

$$U_{iz}(t) = V_{iz}(t) - \epsilon_{iz}(t)$$

\Rightarrow Advantages but two serious problems ...

Model

Elements and assumptions

- $\epsilon_{iz}(t)$ are iid over individuals and time within region, with pdf

$$f(\epsilon) = \frac{1}{\mu} \exp \left[\frac{\epsilon - \eta}{\mu} \right] \exp \left\{ -e^{\frac{\epsilon - \eta}{\mu}} \right\}$$

- Probability that individual i in region z experiences trial at time t is obtained from $F(\epsilon)$ as

$$\begin{aligned} P(y_{iz}(t) = 1) &= P(\epsilon_{iz}(t) \leq V_{iz}(t)) \\ &= 1 - \exp \left\{ -\exp \left\{ \frac{V_{iz}(t) - \eta}{\mu} \right\} \right\}. \end{aligned}$$

- The probability that at least *one* individual tries is

$$\begin{aligned} P(y_z(t) = 1) &= P(\max_i \{ U_{iz}(t) \mid i = 1, \dots, n_z \} \geq 0) \\ &= P(\max_i \{ V_{iz}(t) - \epsilon_{iz}(t) \} \geq 0) \\ &= P(V_z(t) - \min_i \{ \epsilon_{iz}(t) \} \geq 0) \\ &\quad \text{since we have } V_{iz}(t) = V_z(t) \quad \forall i \\ &= P(\min_i \{ \epsilon_{iz}(t) \} \leq V_z(t)) \end{aligned}$$

Model

Elements and assumptions

- Solution is to define region-specific event — probability that unobserved maximal individual's utility exceeds zero is equivalent to probability that observed deterministic utility $V_z(t)$ for the representative individual from region exceeds *minimum* value of all $\epsilon_{iz}(t)$

$$\begin{aligned}\epsilon_{iz}(t) &\sim G(\eta, \mu) \\ \epsilon_z^{min}(t) &= \min_i \{ \epsilon_{iz}, i = 1, \dots, n_z \} \\ &\sim G(\eta - \mu \ln(n_z), \mu).\end{aligned}$$

- So that probability that trial occurs in region z given that it has not yet occurred is obtained as

$$\begin{aligned}P(y_z(t) = 1) &= F(\epsilon_z^{min}(t)) \\ &= 1 - \exp \left\{ - \exp \left\{ \frac{V_z(t) - (\eta - \mu \ln(n_z))}{\mu} \right\} \right\} \\ &= 1 - \exp \left\{ - \exp \{ V_z(t) + \ln(n_z) \} \right\}\end{aligned}$$

Model

Neighborhood effects

- Amend deterministic component of region utility

$$V'_z(t) = V_z(t) + \theta[w_z Y_z(t-1)]$$

- Ensure consistency with rational behavior (BROCK & DURLAF 2001)
- Ensure identification — no “reflection” (MANSKI 1993)
- One candidate for $w_z Y_z(t-1)$ is obtained from first order contiguity

Model

Neighborhood effects

- Lagged expected average choice behavior is a valid representation of social utility (BROCK & DURLAF 2001)
- Identification is possible because relationship between effect and regressors is nonlinear for sufficient variation in neighborhood characteristics
- Implications of reflection
 - True effect if probability of event varies with measure of average probability (behavior) of *exogenous* reference group
 - Contextual effect if probability of trial varies according to the characteristics of the reference group
 - Correlated effect if probability varies due to correlated unobservables

Model

Summary of properties

- *Within Regions.* IID utilities, focus on the *first* trier
- *Across Regions.* Influence flows across exogenously defined groups
- *Rationality.* Model is consistent with RUM, no reflection problem
- *Other.* Gumbel distribution exploited to circumvent lack of individual information; choice of interval length can introduce bias, however complementary log-log model estimates consistent with underlying continuous time process

Empirical Analysis

Data

1. Disaggregate transaction information

- Customer identification code
- Total transaction value
- 382,478 transactions (05/01/97 through 01/31/01)
- 162,618 customers
- 45 discrete time periods

2. Supplementary zip code information

- 29,701 residential zip codes
- Match to census data for control variables
 - (a) Intrinsic characteristics
 - (b) Household economics
 - (c) Local environment

Empirical Analysis

3. CACI retail information

- Zip code summary of retail presence, sales at convenience, drug, supermarket, w/house stores

4. Measure of contagion/neighborhood effect

- **Lagged cumulative effect (LC)**
- Lagged effect (L)

Empirical Analysis

Preliminaries

- Individuals
 - Average order value \$ 51.53 (SD = \$50.99) [supermarket \$29.80 (SD = \$29.18)]
- Regions
 - 29,701 residential zip codes with 369,146 orders and 156,069 customers
 - 1,508 non-residential zip codes with 25,123 orders (eliminated)
 - Trial penetration is approximately 60 percent by 01/31/01
 - Average number of contiguous neighbors = 5.61 (SD = 2.30), some “islands”
- National Space-Time
- Local Space-Time

Empirical Analysis

Results: Initial evidence

- Neighborhood effect only models
 - Show significant effects for all formulations
 - Suggest cumulative approach is best
 - Support distributional assumptions, model structure — coefficient on $\ln(n_z)$ very stable
- But ... is the effect “real”
 - Unobserved common traits
 - Unobserved heterogeneity
 - Endogeneity
 - (Unobserved correlated process)

Empirical Analysis

Results: Further evidence

- Expand formulation with
 - Non-parametric time-dependent baseline hazard (heterogeneity)
 - Observed heterogeneity across regions
 - State-level fixed effects (unobserved common traits)
 - State-level mean observables as instruments
 - Internet access, random effect
- After introduction of controls (120 variables), θ
 - Diminishes in magnitude
 - Remains statistically significant
 - Is second most important variable (Wald χ^2 and standardized coefficients)
 - Holds under alternative formulations for n_z and $w_z Y_z(t - 1)$

Empirical Analysis

Results: Substantive implications

- Approximately nineteen percent increase in baseline hazard

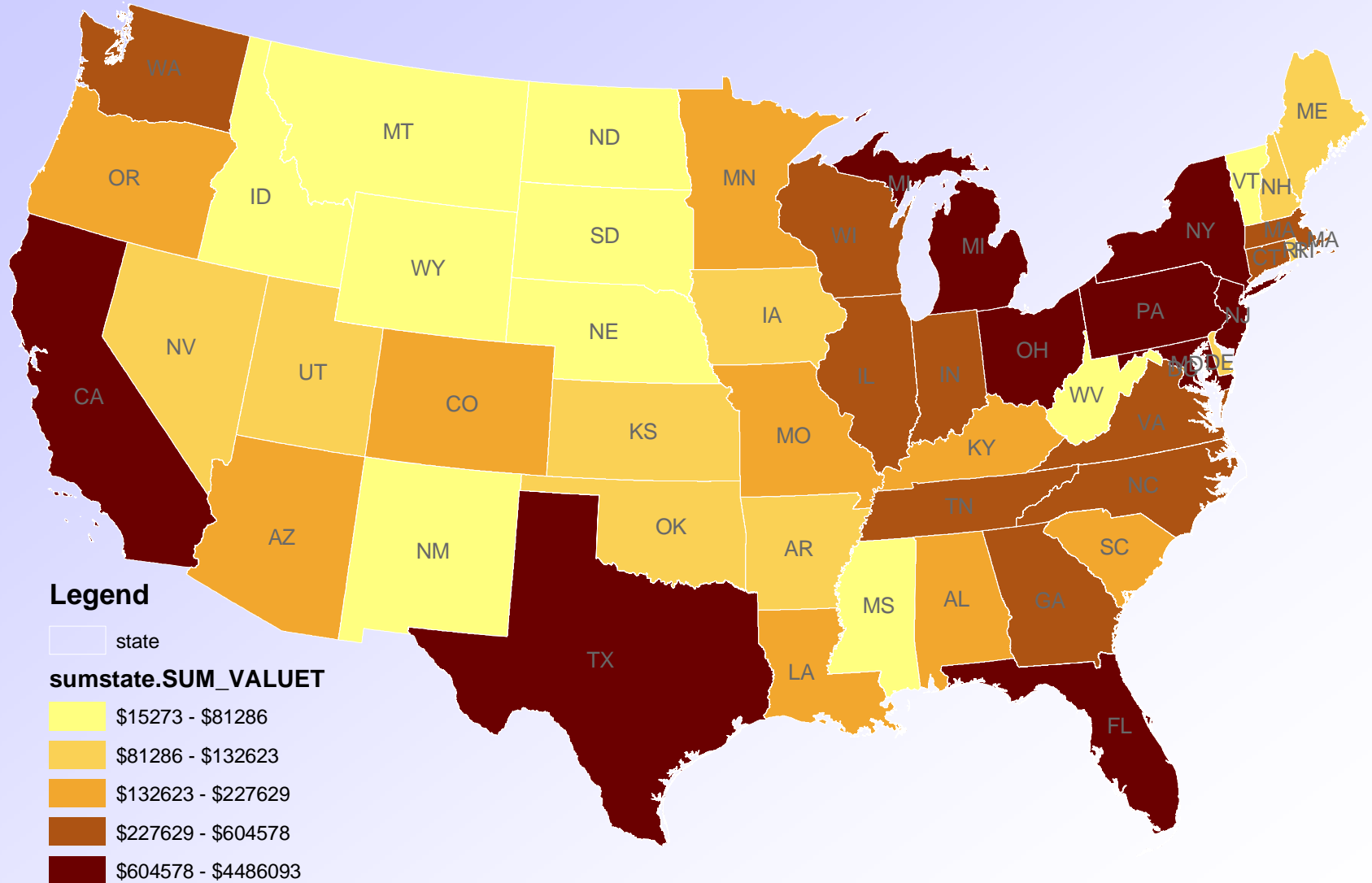
⇒ Marginal effect of zero to 20,000 neighbors trying, increases focal zip code probability from about 2.7% to 14.0

- Empirical findings
 - (1) Household Characteristics
 - (2) Household Economics
 - (3) Local environment
 - (4) Access to Retail Services

Conclusions

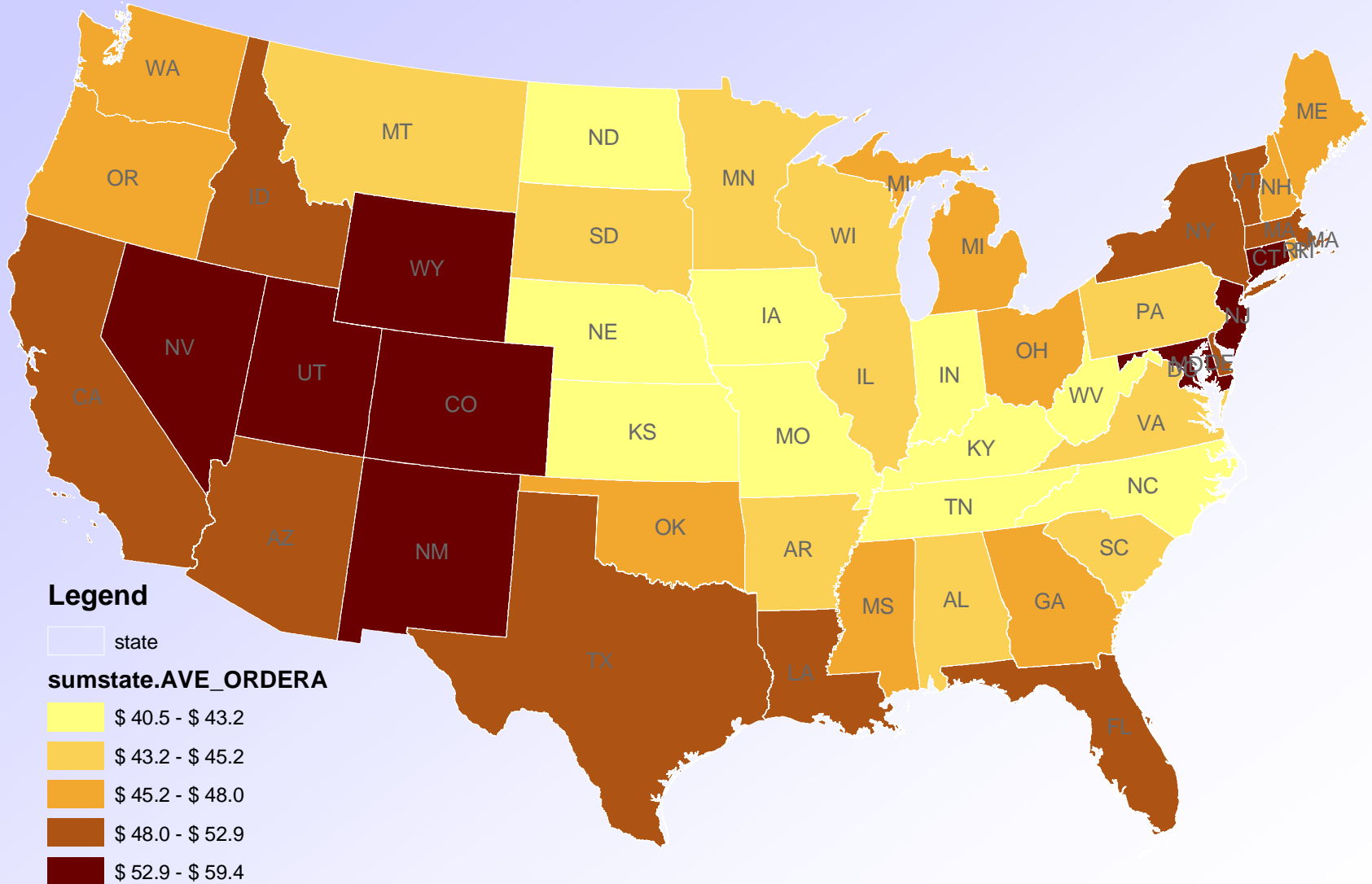
- Evidence suggests that neighborhood effects
 - operate on Internet \Rightarrow social observation/exchange grounded in proximity is important
 - (dissipate when individuals have own information?)
 - (could be exploited through judicious seeding)?
- Discrete time hazard model for continuous time process with an unobserved risk set can be derived to link individuals and regions
- Future research
 - Affiliation based on “socio-demographic proximity” (working paper)
 - Preference minorities (in progress)

Total Order Value



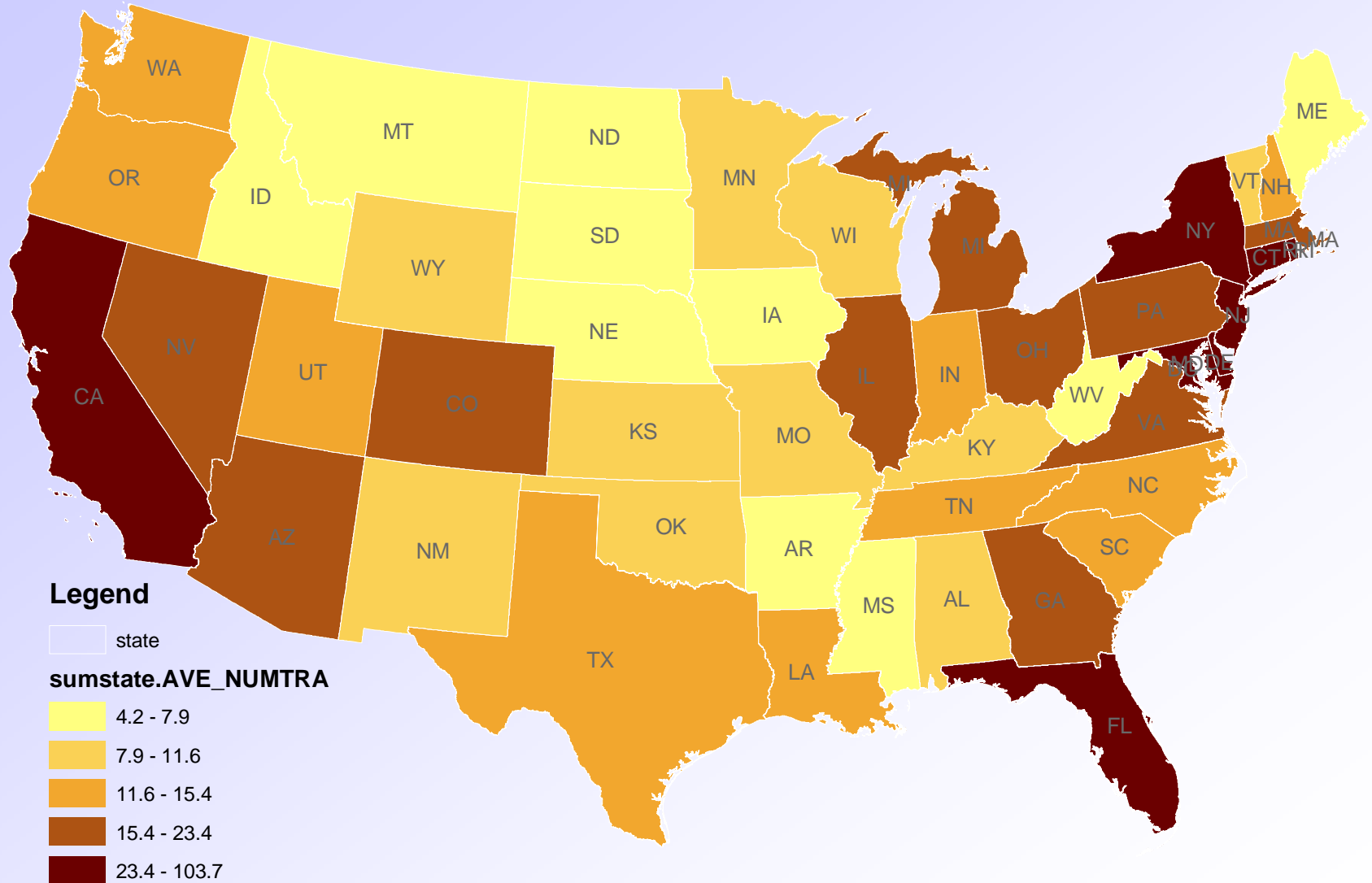
FTC May 24, 2007

Average Order Value per Customer



FTC May 24, 2007

Average Number of Transactions



FTC May 24, 2007